

▼ DOWNLOADING NECESSARY NLP TOOLKIT

```
import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('stopwords')
nltk.download('omw-1.4')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
True
```

▼ IMPORTING LIBRARIES

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import re

import nltk
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
from wordcloud import WordCloud
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from imblearn.over_sampling import SMOTE

#IMPORTING ML MODELS
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import MultinomialNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

▼ LOADING DATASET

```
df=pd.read_csv('/content/train.csv')
df
```

	id	label	tweet	
0	1	0	@user when a father is dysfunctional and is s...	
1	2	0	@user @user thanks for #lyft credit i can't us...	
2	3	0	bihday your majesty	
3	4	0	#model i love u take with u all the time in ...	
4	5	0	factsguide: society now #motivation	
...	
31957	31958	0	ate @user isz that youuu?ð ð ð ð ð ð...	
31958	31959	0	to see nina turner on the airwaves trying to...	
31959	31960	0	listening to sad songs on a monday morning otw...	
31960	31961	1	@user #sikh #temple vandalised in in #calgary,...	
31961	31962	0	thank you @user for you follow	

31962 rows × 3 columns

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31962 entries, 0 to 31961
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0    id      31962 non-null   int64
1   label   31962 non-null   int64
2   tweet    31962 non-null   object
dtypes: int64(2), object(1)
memory usage: 749.2+ KB
```

▼ DROPPING UNCESSARY COLUMNS

```
df.drop(['id'],inplace=True,axis=1)
```

▼ DROPING DUPLICATES VALUE

```
df.drop_duplicates(keep="first")
```

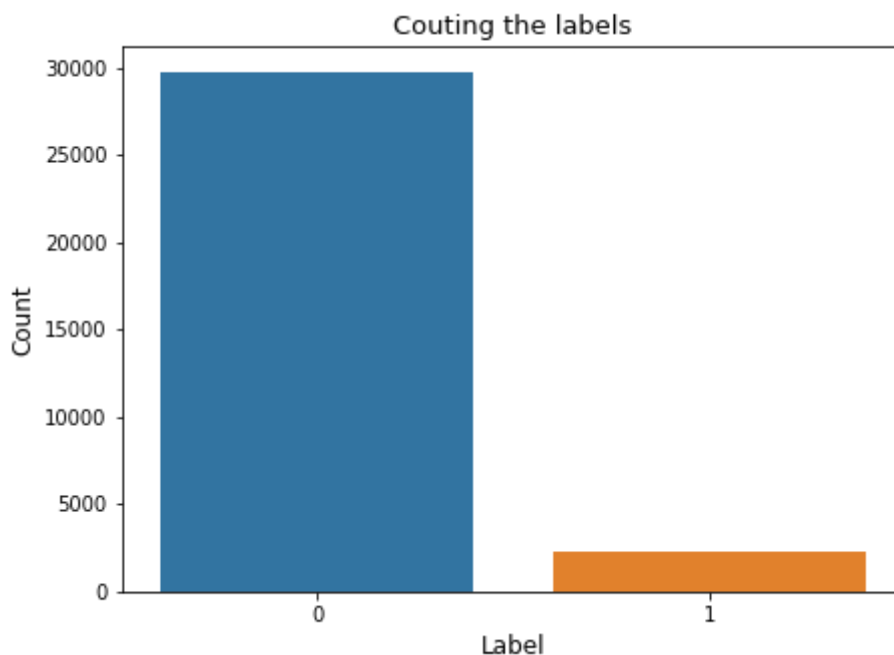
	label	tweet
0	0	@user when a father is dysfunctional and is s...
1	0	@user @user thanks for #lyft credit i can't us...
2	0	bihday your majesty
3	0	#model i love u take with u all the time in ...
4	0	factsguide: society now #motivation
...
31956	0	off fishing tomorrow @user carnt wait first ti...
31957	0	ate @user isz that youuu?ö ö ö ö ö ö...
31958	0	to see nina turner on the airwaves trying to...
31959	0	listening to sad songs on a monday morning otw...
31961	0	thank you @user for you follow

29530 rows × 2 columns

▼ DATA AUGUMENTATION USING NLPAUG

```
plt.figure(figsize=(7, 5))
plt.title('Couting the labels', fontsize=13)
sns.countplot(data=df, x='label')
plt.xlabel('Label', fontsize=12)
plt.ylabel('Count', fontsize=12)
```

```
Text(0, 0.5, 'Count')
```



```
!pip install nlpaug
!pip install textaugment
!pip install transformers
```

```
Requirement already satisfied: requests>=2.22.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: textaugment in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: googletrans in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: gensim in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: scipy>=0.18.1 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: httpx==0.13.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: chardet==3.* in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: sniffio in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: hstspreload in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: httpcore==0.9.* in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: rfc3986<2,>=1.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: idna==2.* in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: h11<0.10,>=0.8 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: h2==3.* in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: hpack<4,>=3.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: hyperframe<6,>=5.2.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.9.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: tokenizers!=0.11.3,<0.13,>=0.11.1 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from transformers==4.2.0)
```

Requirement already satisfied: certifi<2021.4.1 in /usr/local/lib/python3.7/dist-packages
 Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
 Requirement already satisfied: urllib3<1.25.0, >=1.25.1, <1.26, >=1.21.1 in /usr/local

```
import nlpaug.augmenter.char as nac
import nlpaug.augmenter.word as naw
import nlpaug.augmenter.sentence as nas
import nlpaug.flow as nafc
```

```
data_df=df.copy()
```

```
aug_texts = []
minority_data = data_df[data_df['label'] == 1]
aug = naw.SynonymAug(aug_src='wordnet')
```

```
texts = minority_data['tweet'].tolist()
```

```
for text in texts:
    augmented_texts = aug.augment(text, n=12)
```

```
    for augmented in augmented_texts:
        aug_texts.append(augmented)
```

```
print(len(aug_texts))
```

```
26904
```

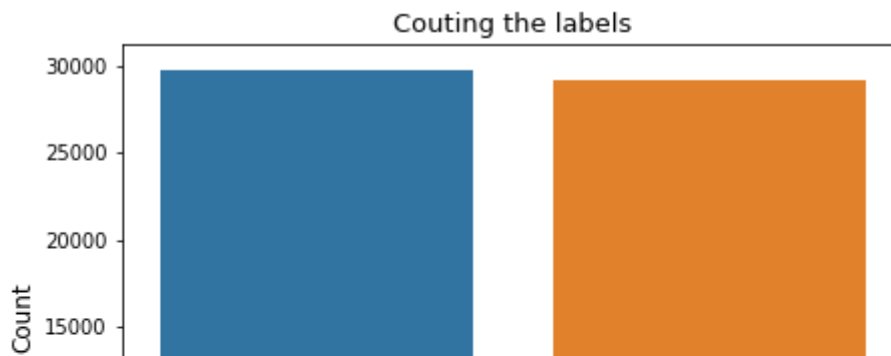
```
temp = pd.DataFrame({
    'tweet': aug_texts
})
```

```
temp['label']=1
```

```
data_resampled_nlpaug = pd.concat([data_df, temp], axis=0)
data_resampled_nlpaug = data_resampled_nlpaug.reset_index()
data_resampled_nlpaug = data_resampled_nlpaug.drop(columns=['index'])
del temp, minority_data
```

```
plt.figure(figsize=(7, 5))
plt.title('Counting the labels', fontsize=13)
sns.countplot(data=data_resampled_nlpaug, x='label')
plt.xlabel('Label', fontsize=12)
plt.ylabel('Count', fontsize=12)
```

Text(0, 0.5, 'Count')



▼ CLEANING TWEET BY MAKING NEW COLUMN

```

from html.parser import HTMLParser
html_parser=HTMLParser()

df['clean_tweet']=df['tweet'].apply(lambda x:html_parser.unescape(x))
df.head(10)

```

	label	tweet	
0	0	@user when a father is dysfunctional and is s...	@user when a father is dy
1	0	@user @user thanks for #lyft credit i can't us...	@user @user thanks for #
2	0	bihday your majesty	
3	0	#model i love u take with u all the time in ...	#model i love u take w
4	0	factsguide: society now #motivation	factsguide: soc
5	0	[2/2] huge fan fare and big talking before the...	[2/2] huge fan fare and bi
6	0	@user camping tomorrow @user @user @user @use...	@user camping tomorrow
7	0	the next school year is the year for exams.ð ...	the next school year is the y
8	0	we won!!! love the land!!! #allin #cavs #champ...	we won!!! love the land!!! #
9	0	@user @user welcome here ! i'm it's so #gr...	@user @user welcome

▼ REMOVING @USER FROM ALL TWEETS

```

def remove_pattern(input_txt,pattern):
    r = re.findall(pattern,input_txt)
    for i in r:
        input_txt=re.sub(i, '', input_txt)
    return input_txt

df['clean_tweet']=np.vectorize(remove_pattern)(df['tweet'],"@[\w]*")
df['clean_tweet'][:3]

```

```
0      when a father is dysfunctional and is so sel...
1      thanks for #lyft credit i can't use cause th...
2      bihday your majesty
Name: clean_tweet, dtype: object
```

df

	label	tweet	
0	0	@user when a father is dysfunctional and is s...	when a father is dysfun
1	0	@user @user thanks for #lyft credit i can't us...	thanks for #lyft credit
2	0	bihday your majesty	
3	0	#model i love u take with u all the time in ...	#model i love u take w
4	0	factsguide: society now #motivation	factsguide: soc
...	
31957	0	ate @user isz that youuu? ø ø ø ø ø ø...	ø ø ø ø
31958	0	to see nina turner on the airwaves trying to...	to see nina turner on th
31959	0	listening to sad songs on a monday morning otw...	listening to sad songs on a m
31960	1	@user #sikh #temple vandalised in in #calgary,...	#sikh #temple vandalised i
31961	0	thank you @user for you follow	th

31962 rows × 3 columns

➤ CONVERTING ALL THE TWEETS INTO LOWERCASE

```
df['clean_tweet'] = df['clean_tweet'].apply(lambda x: x.lower())
df.head(10)
```

label		tweet	
0	0	@user when a father is dysfunctional and is s...	when a father is dysfunctiona

REMOVING PUNCTUATIONS

```
df['clean_tweet']=df['clean_tweet'].apply(lambda x: re.sub(r'^\w\s',' ',x))

[2/21 huge fan fare and big talking before the

#REMOVING INTEGERS WITH SPACE
```

```
df['clean_tweet']=df['clean_tweet'].apply(lambda x: re.sub(r'^a-zA-Z',' ',x))

we would love the remain from #cave #example
```

REPLACING SPEACIAL CHARACTERS WITH SPACE

```
df['clean_tweet'] = df['clean_tweet'].apply(lambda x: re.sub(r'^a-zA-Z0-9',' ',x))

df
```

label		tweet	
0	0	@user when a father is dysfunctional and is s...	when a father is dysfunc
1	0	@user @user thanks for #lyft credit i can't us...	thanks for lyft credit
2	0	bihday your majesty	
3	0	#model i love u take with u all the time in ...	model i love u take w
4	0	factsguide: society now #motivation	factsguide s
...	
31957	0	ate @user isz that youuu?	
31958	0	to see nina turner on the airwaves trying to...	to see nina turner on th
31959	0	listening to sad songs on a monday morning otw...	listening to sad songs c
31960	1	@user #sikh #temple vandalised in in #calgary,...	sikh temple vandalise
31961	0	thank you @user for you follow	the

SPELLING CORRECTION USING TEXTBLOB LIBRARY

```
from textblob import TextBlob
```



```
text=df['clean_tweet'][0:10].apply(lambda x: str(TextBlob(x).correct()))
text
```

```
0      when a father is dysfunctional and is so sel...
1      thanks for left credit i can t use cause th...
2      midday your majesty
3      model i love u take with u all the time in ...
4      factsguide society now motivation
5      huge fan fare and big talking before the...
6      camping tomorrow dandy
7      the next school year is the year for exam ...
8      we won love the land allen caps champ...
9      welcome here i m it s so gr
Name: clean_tweet, dtype: object
```

▼ TOKENIZATIONS

```
df['tokens']=df['clean_tweet'].apply(lambda x:word_tokenize(x))
```

df

	label	tweet	clean_tweet	
0	0	@user when a father is dysfunctional and is s...	when a father is dysfunctional and is so sel...	dys
1	0	@user @user thanks for #lyft credit i can't us...	thanks for lyft credit i can t use cause th...	[thanks,
2	0	bihday your majesty	bihday your majesty	[l
3	0	#model i love u take with u all the time in ...	model i love u take with u all the time in ...	[model, i,
4	0	factsguide: society now #motivation	factsguide society now motivation	[fac
...	
31957	0	ate @user isz that youuu? ø ø ø ø ø ø...	ate isz that youuu ...	
31958	0	to see nina turner on the airwaves trying to...	to see nina turner on the airwaves trying to...	[to, see
31959	0	listening to sad songs on a monday	listening to sad songs on a	[listenin

▼ REMOVING STOPWORDS

```
stopwords=stopwords.words('english')
stopwords
'same',
```

```
'so',  
'than',  
'too',  
'very',  
's',  
't',  
'can',  
'will',  
'just',  
'don',  
"don't",  
'should',  
"should've",  
'now',  
'd',  
'll',  
'm',  
'o',  
're',  
've',  
'y',  
'ain',  
'aren',  
"aren't",  
'couldn',  
"couldn't",  
'didn',  
"didn't",  
'doesn',  
"doesn't",  
'hadn',  
"hadn't",  
'hasn',  
"hasn't",  
'haven',  
"haven't",  
'isn',  
"isn't",  
'ma',  
'mightn',  
"mightn't",  
'mustn',  
"mustn't",  
'needn',  
"needn't",  
'shan',  
"shan't",  
'shouldn',  
"shouldn't",  
'wasn',  
"wasn't",  
'weren',  
"weren't",  
'won',  
"won't",  
'wouldn',  
"wouldn't"]
```

```
df['removed_stopword']=df['tokens'].apply(lambda x: [word for word in x if not word in sto
```

df

	label	tweet	clean_tweet	tokens
0	0	@user when a father is dysfunctional and is s...	when a father is dysfunctional and is so sel...	[when, a, father, is, dysfunctional, and, is, ...]
1	0	@user @user thanks for #lyft credit i can't us...	thanks for lyft credit i can t use cause th...	[thanks, for, lyft, credit, i, can, t, use, ca...]
2	0	bihday your majesty	bihday your majesty	[bihday, your, majesty]
3	0	#model i love u take with u all the time in ...	model i love u take with u all the time in ...	[model, i, love, u, take, with, u, all, the, t...]
4	0	factsguide: society now #motivation	factsguide society now motivation	[factsguide, society, now, motivation]
...
31957	0	ate @user isz that youuu? ö ö ö ö ö ö...	ate isz that youuu ...	[ate, isz, that, youuu]
31958	0	to see nina turner on the airwaves trying to...	to see nina turner on the airwaves trying to...	[to, see, nina, turner, on, the, airwaves, try...]
31959	0	listening to sad songs on a monday morning otw	listening to sad songs on a monday morning	[listening, to, sad, songs, on, a, ...]

▼ USING STEMMING TECHNIQUES

```
stem=PorterStemmer()
df['stem']=df['removed_stopword'].apply(lambda x:' '.join([stem.stem(i) for i in x]))
```

```
df['stem']
```

```
0      father dysfunct selfish drag kid dysfunct run
1      thank lyft credit use caus offer wheelchair va...
2                               bihday majesti
3                model love u take u time ur
4                factsguid societi motiv
...
31957                                ate isz youuu
31958      see nina turner airwav tri wrap mantl genuin h...
31959                listen sad song monday morn otw work sad
31960                sikh templ vandalis calgari wso condemn act
31961                                thank follow
Name: stem, Length: 31962, dtype: object
```

▼

CONVERTING WORDS INTO ITS BASE FORM USING LEMMITIZATION

```
lemma=WordNetLemmatizer()

df['lemmatized']=df['removed_stopword'].apply(lambda x:' '.join([lemma.lemmatize(i) for i
in x.split(' ')])

df.head(10)
```

	label	tweet	clean_tweet	tokens	removed_stopword	s
0	0	@user when a father is dysfunctional and is s...	when a father is dysfunctional and is so sel...	[when, a, father, is, dysfunctional, and, is, ...	[father, dysfunctional, selfish, drags, kids, ...	father dysfu selfish drag dysfunct
1	0	@user @user thanks for #lyft credit i can't us...	thanks for lyft credit i can t use cause th...	[thanks, for, lyft, credit, i, can, t, use, ca...	[thanks, lyft, credit, use, cause, offer, whee...	thank lyft cr use caus c wheelcl v
2	0	bihday your majesty	bihday your majesty	[bihday, your, majesty]	[bihday, majesty]	bihday maj
3	0	#model i love u take with u all the time in ...	model i love u take with u all the time in ...	[model, i, love, u, take, with, u, all, the, t...	[model, love, u, take, u, time, ur]	model lov take u time
4	0	factsguide: society now #motivation	factsguide society now motivation	[factsguide, society, now, motivation]	[factsguide, society, motivation]	factsc societi m
5	0	[2/2] huge fan fare and big talking before the...	huge fan fare and big talking before the...	[huge, fan, fare, and, big, talking, before, t...	[huge, fan, fare, big, talking, leave, chaos, ...	huge fan f big talk l chao disput g
6	0	@user camping tomorrow @user @user @user @use...	camping tomorrow danny	[camping, tomorrow, danny]	[camping, tomorrow, danny]	ca tomor da

▼

VISUALIZING THE DATA USING LIBRARY CALLED WORDCLOUD

```
positive_tweets=' '.join(df[df['label']==0]['clean_tweet'])
negative_tweets=' '.join(df[df['label']==1]['clean_tweet'])
```

```
plt.figure(figsize=(7,7))
plt.imshow(wc)
plt.axis("off")
plt.show()
```



```
plt.figure(figsize=(7,7))
plt.imshow(wc)
plt.axis("off")
plt.show()
```



Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/>
Requirement already satisfied: emoji in /usr/local/lib/python3.7/dist-packages (2.1.0)

df

index	label	tweet	clean_tweet	
0	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	when a father is dysfunctional and is so selfish he drags his kids into his dysfunction run	when,a,father,is,dysfunctional,and,is,so,selfish,
1	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	thanks for lyft credit i can t use cause they don t offer wheelchair vans in pdx disapointed getthanked	thanks,for,lyft,credit,i,can,t,use,cause,they,don,
2	0	bihday your majesty	bihday your majesty	bihday,your,majesty
3	0	#model i love u take with u all the time in ur ð ±!!! ð ð ð ð ð ð ð	model i love u take with u all the time in ur	model,i,love,u,take,with,u,all,the,time,in,ur
4	0	factsguide: society now #motivation	factsguide society now motivation	factsguide,society,now,motivation
5	0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo	huge fan fare and big talking before they leave chaos and pay disputes when they get there allshowandnogo	huge,fan,fare,and,big,talking,before,they,leave,
6	0	@user camping tomorrow @user @user @user @user @user dannyâ	camping tomorrow danny	camping,tomorrow,danny
7	0	the next school year is the year for exams.ð — can't think about that ð #school #exams #hate #imagine #actorslife #revolutionschool #girl	the next school year is the year for exams can t think about that school exams hate imagine actorslife revolutionschool girl	the,next,school,year,is,the,year,for,exams,can,t
8	0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers â	we won love the land allin cavs champions cleveland clevelandcavaliers	we,won,love,the,land,allin,cavs,champions,clev
9	0	@user @user welcome here ! i'm it's so #gr8 !	welcome here i m it s so gr	welcome,here,i,m,it,s,so,gr
10	0	â #ireland consumer price index (mom) climbed from previous 0.2% to 0.5% in may #blog #silver #gold #forex	ireland consumer price index mom climbed from previous to in may blog silver gold forex	ireland,consumer,price,index,mom,climbed,for
		we are so selfish.	we are so selfish	

11	0	#orlando #standwithorlando #pulseshooting #orlandoshooting #biggerproblems #selfish #heabreaking #values #love #	orlando standwithorlando pulseshooting orlandoshooting biggerproblems selfish heabreaking values love	we,are,so,selfish,orlando,standwithorlando,puls
12	0	i get to see my daddy today!! #80days #gettingfed	i get to see my daddy today days gettingfed	i,get,to,see,my,daddy,today,days,gettingfed
13	1	@user #cnn calls #michigan middle school 'build the wall' chant "	cnn calls michigan middle school build the wall	cnn,calls,michigan,middle,school,build,the,wall,

▼ SPLITTING THE DATA INTO X AND Y VARIABLES

14	1	#seashepherd	seashepherd	no.comment.in.australia.opkillinbav.seasheph
----	---	--------------	-------------	--

```
x=df['clean_tweet']
y=df['label']
```

		ouch i junior is	ouch i junior is	
--	--	------------------	------------------	--

▼ TRAINING AND TESTING PHASE

15	0	parish priest	parish priest	parish priest,parish priest,parish priest,parish priest,parish priest
----	---	---------------	---------------	---

```
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_state=11)
```

		agree	agree	
--	--	-------	-------	--

▼ COUNTER VECTORS METHOD

		as we all know	as we all know	
--	--	----------------	----------------	--

```
vectorizer=CountVectorizer()
vectorizer.fit(xtrain)

xtrain_countvc=vectorizer.transform(xtrain)
xtest_countvc=vectorizer.transform(xtest)

print(xtrain_countvc.toarray())
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

22	0	who's it's the #weekend?	who's it's the #weekend?	product,of,the,day,happy,man,wine,tool,who,s,i
----	---	--------------------------	--------------------------	--

▼ "TF AND IDF" METHODS

23	1	says i am a . prove it	prove it lumpy	lumpy,says,i,am,a,prove,it,lumpy
----	---	------------------------	----------------	----------------------------------

```
tfidf=TfidfVectorizer()
tfidf.fit(xtrain)

xtrain_tfidf=tfidf.transform(xtrain)
xtest_tfidf=tfidf.transform(xtest)

print(xtrain_tfidf.toarray())

[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

▼ USING LOGISTICS REGRESSION

```
lr=LogisticRegression()
lr.fit(xtrain_countvc,ytrain)
```

```
y_pred=lr.predict(xtest_countvc)
lr.score(xtest_countvc,ytest)
```

```
0.9593283971217019
```

```
y_test_pred=lr.predict(xtest_tfidf)
lr.score(xtest_tfidf,y_test_pred)
```

```
📄 1.0
```

```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
```

```
print("Accuracy of the model",accuracy_score(y_pred,ytest))
print("CONFUSION-MATRIX",confusion_matrix(y_pred,ytest))
print("Classification reprot",classification_report(y_pred,ytest))
```

```
'
precision    recall  f1-score   support\n\n
9183\n      1      0.51      0.87      0.65      406\n\n
0 96      9589\n  macro avg      0 75      0 92      0 81      9589\nweighted avg
```

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 8:35 PM

